

Salivary Proteome Wiki: A Collaborative Environment to Share and Annotate Proteomics Data

W Lau¹, L Shum², C Johnson¹

¹Division of Computational Bioscience, Center for Information Technology,

²Division of Extramural Research, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland 20892, USA

Introduction

With the objective of enhancing our understanding of oral health and disease pathogenesis, researchers have recently completed a catalog of 1,116 proteins in human saliva using the shotgun proteomics approach (1).

The catalog that has been put together is an important first step towards the discovery and characterization of saliva biomarkers for various diseases. However, much effort is still required to enrich and refine the catalog, especially those proteins with relatively weak experimental evidence. To realize this goal, active participation from the research community at large is crucial.

Methods

The Salivary Proteome Wiki (SPW) is developed to as a central repository for researchers to store peptide and protein identification results. Contents are organized around the proteins that have been identified, integrating data from different research groups to ensure accurate characterization of the proteins. The system is built upon the MediaWiki software utilized in Wikipedia, with additional semantic plugins to help organize and evaluate the data. A text mining tool, AKS², has also been incorporated. Unlike other databases, the wiki encourages maximum level of interactivity and collaboration, where users have the ability to propose a wide series of changes on any data, ranging from adding new annotations to creating new data types to changing the way that information are displayed.

Results

Many proteome projects, such as the Human Salivary Proteome Project, involve researchers from across geographical locations. In addition to issues with data integration, working together on a day-to-day basis is not easy. SPW provides a collaborative model for exchanging ideas, sharing experimental results, and discovering new knowledge within a single platform.

Currently, there are more than 170 experiments, 2300 protein identifications, and 175,000 peptide hits stored in SPW as wiki pages. Pages are linked together through semantic annotations. These annotations are also used to create summary charts that offer a high level view of the data. In addition, users have access to a number of Web services (2), such as BLAST, to carry out further

analysis. They can annotate the data using controlled vocabulary terms or free-text. Contributions can also be made by participating in discussions on associated talk pages.

As increasing volume of data are generated from high-throughput experiments, the ability to exploit all available resources within the community becomes more appealing than ever. Using Web 2.0 and Semantic Web technologies, this vision is one step closer to reality.

Innovative aspects

- Collaborative content development model with discussion, annotation, data import/export, and curation features
- Semantic markups to support complex *ad hoc* querying and machine-based reasoning
- Built-in tools to facilitate storage and analysis of proteomics data

References

- (1) P Denny *et al*, The proteomes of human parotid and submandibular/sublingual gland salivas collected as the ductal secretions; J. Proteome Res. 2008, 7:1994-2006.
- (2) A Labarga *et al*, Web Services at the European Bioinformatics Institute; Nucleic Acids Res. 2007, 35:W6-11.



Figure 1. Examples of figures generated in the Salivary Proteome Wiki using semantic annotations on the data to provide a real-time summary of the information stored in the system.